How to Detect Linear Dependence on the Copula Level?

Vladik Kreinovich[1], **Hung T. Nguyen**[2,3], and **Songsak Sriboonchitta**[3]

[1]Department of Computer Science, University of Texas at El Paso
500 W. University, El Paso, TX 79968, USA, vladik@utep.edu

[2] Department of Mathematical Sciences, New Mexico State University
Las Cruces, New Mexico 88003, USA, hunguyen@nmsu.edu

[3]Department of Economics, Chiang Mai University
Chiang Mai, Thailand, songsak@econ.chiangmai.ac.th

1. Linear Dependencies Are Ubiquitous

- Dependencies between quantities are often described by smooth (even analytical) functions $y = f(x_1, \ldots, x_n)$:

$$y = f(x^{(0)}) + \sum_{i=1}^{n} c_i \cdot (x_i - x_i^{(0)}) + \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} \cdot (x_i - x_i^{(0)}) \cdot (x_j - x_j^{(0)}) + \ldots$$

- For values x_i close to $x_i^{(0)}$, we can safely ignore terms which are quadratic in $x_i - x_i^{(0)}$ (or of higher order):

$$y \approx f(x^{(0)}) + \sum_{i=1}^{n} c_i \cdot (x_i - x_i^{(0)}).$$

- Linear dependencies often extend beyond local.
- Linear dependencies make computations easier; e.g.:
  - systems of linear equations can be efficiently solved,
  - systems of non-linear equations are NP-hard.
- It is thus important to detect linear dependencies.

2. How Linear Dependence is Detected Now

- *Exact* linear dependence can be detected by whether the corr. system of linear equations has a solution:

$$y^{(k)} = f(x^{(0)}) + \sum_{i=1}^{n} c_i \cdot \left(x_i^{(k)} - x_i^{(0)}\right), \quad k = 1, \ldots, K.$$

- There exist efficient algorithms for checking solvability of such a linear system.
- *Approximate* linearity can be gauged by computing the Pearson's correlation coefficient $r(F) = \dfrac{C_{XY}}{\sigma_X \cdot \sigma_Y}$.
- In the case of an exact linear dependence $Y = c_0 + c_1 \cdot X$, we have $r(F) = 1$ if $c_1 > 0$ and $r_F = -1$ if $c_1 < 0$.
- The square $R^2 = (r(F))^2$ is a measure of fit with the linear model: the closer R^2 to 1, the better the fit.

3. Detecting Linear Dependence Based on a Copula: 1st Problem

- A joint distribution of X and Y can be described by the cdf $F(x,y) \stackrel{\text{def}}{=} \text{Prob}\,(X \le x\,\&\,Y \le y)$.
- This description depends on the units in which we describe x and y: m to cm, log scale, etc.
- A unit-independent description is known as a *copula*:

  $C(a,b) \stackrel{\text{def}}{=} F(x,y)$ for x,y s.t. $a = F_X(x)$ and $b = F_Y(y)$.

- Here, $F(x,y) = C(F_X(x), F_Y(y))$, i.e.,

$$C(a,b) = F(F_X^{-1}(a), F_Y^{-1}(b)).$$

- How can we detect linear dependence between the quantities x and y based only on the copula $C(a,b)$?

4. Detecting Linear Dependence Based on a Copula: Main Idea and the Resulting Definition

- For a given copula $C(a,b)$, possible cdfs are of the form
$$F(x,y)=C(F_X(x),F_Y(y)).$$
- A dependence is linear if $r=\pm 1$ for some marginals $F_X(x)$ and $F_Y(y)$.
- So, we define measures of linearity as
$$L^- \stackrel{\text{def}}{=} \min_{F_X(x),F_Y(y)} r(C(F_X(x),F_Y(y));$$
$$L^+ \stackrel{\text{def}}{=} \max_{F_X(x),F_Y(y)} r(C(F_X(x),F_Y(y)).$$
- The values L^- and L^+ depend only on the copula.
- As a measure of fit, we can use $M=\max((L^-)^2,(L^+)^2)$.
- $M=1$ if and only if the dependence is linear for some marginals.

5. How to Actually Compute L^- and L^+: Towards an Algorithm

- For any X and Y corr. to the copula, all others are obtained by re-scaling $X' = A(X)$ and $Y' = B(Y)$.
- Thus, L^- is min and L^+ is max of the expression $L \stackrel{\text{def}}{=} r(C(A(x), B(y))$ over all possible $A(x)$ and $B(y)$.
- Min, max are attained when $\dfrac{\delta L}{\delta A(x)} = \dfrac{\delta L}{\delta B(y)} = 0$, so:

$$A(x) = a_1 + a_2 \cdot E[B(Y) \,|\, X = x]; B(y) = b_1 + b_2 \cdot E[A(X) \,|\, Y = y].$$

- W.l.o.g., we can take $A(0) = B(0)$, $A(1) = B(1) = 1$; then:

$$A(x) = \frac{E[B(Y)) \,|\, X = x] - E[B(Y) \,|\, X = 0]}{E[B(Y) \,|\, X = 1] - E[B(Y) \,|\, X = 0]};$$

$$B(y) = \frac{E[A(X) \,|\, Y = y] - E[A(X) \,|\, Y = 0]}{E[A(X) \,|\, Y = 1] - E[A(X) \,|\, Y = 0]}.$$

6. Resulting Algorithm

- We start with some initial functions $A^{(0)}(x)$ and $B^{(0)}(y)$.
- For example, we can take $A^{(0)}(x) = x$ and $B^{(0)}(y) = y$.
- Once we know $A^{(k)}(x)$ and $B^{(k)}(y)$, we compute:

$$A^{(k+1)}(x) = \frac{E[B^{(k)}(Y)) \mid X = x] - E[B^{(k)}(Y) \mid X = 0]}{E[B^{(k)}(Y) \mid X = 1] - E[B^{(k)}(Y) \mid X = 0]},$$

$$B^{(k+1)}(y) = \frac{E[A^{(k)}(X) \mid Y = y] - E[A^{(k)}(X) \mid Y = 0]}{E[A^{(k)}(X) \mid Y = 1] - E[A^{(k)}(X) \mid Y = 0]}.$$

- We stop when

$$|A^{(k+1)}(x) - A^{(k+1)}(x)| \le \varepsilon; \ \ |B^{(k+1)}(y) - B^{(k+1)}(y)| \le \varepsilon.$$

- We then compute $L^{\pm} = r(A^{(k+1)}(x), B^{(k+1)}(y))$.
- *Testing:* for jointly distributed Gaussian variables, this indeed leads to Pearson's correlation $r(F)$.

7. Two Important Mathematical Subtleties

1. X is not well correlated with $Y = X$ when $X \geq 0$ and $Y = Z \neq X$ for $X < 0$.

- However, $L(A(X), B(Y)) = 1$ when $A(x) = x$ and $B(y) = y$ for $x, y \geq 0$ and $A(x) = B(y) = 0$ else.
- We thus need to make sure that $A(x)$ and $B(y)$ are never constant.
- So, for some $\delta > 0$, we require that $A'(x) \geq \delta$ and $B'(y) \geq \delta$ for all x and y.

2. Max, min are always attained only on a compact set.

- Due to Ascoli-Arzela theorem, compactness means that functions should be uniformly continuous.
- So, we select $M > 0$ and require that $A'(x) \leq M$ and $B'(y) \leq M$ for all x and y.

Under these requirements, the definition works.

8. Heavy-Tailed Distribution: 2nd Problem

- Pearson's correlation $r(F) = \dfrac{C_{XY}}{\sigma_X \cdot \sigma_Y}$ assumes that the marginal distributions have finite variance.
- In reality, however, many econometric-related distributions are *heavy-tailed*, with infinite variance.
- In the 1960s, B. Mandelbrot showed that large-scale fluctuations have pdf $\rho(y) = A \cdot y^{-\alpha}$, with $\alpha \approx 2.7$.
- For this distribution, variance is infinite.
- Since then, similar heavy-tailed distributions have been empirically found in other financial situations.
- We thus need to extend our definitions to the heavy-tailed case.

9. Utility: Reminder

- People's behavior is determined by their preferences.
- A standard way to describe preferences of a decision maker is to use the notion of *utility* u.
- A user prefers an alternative for which the expected value $\sum\limits_{i=1}^{n} p_i \cdot u_i$ of the utility is the largest possible.
- Alternative, we can say that the expected value $\sum\limits_{i=1}^{n} p_i \cdot U_i$ of the *disutility* $U \stackrel{\text{def}}{=} -u$ is the smallest.
- For a random variable, we select an estimate m that minimizes expected disutility:

$$d_U(X) \stackrel{\text{def}}{=} \min_m E[U(Y-m)] = \min_m \int U(y-m) \cdot \rho(y)\, dy.$$

10. Case of Approximate Linear Dependence

- If we only know the variable Y, then we select the estimate m minimizing disutility:

$$d_U(Y) \stackrel{\text{def}}{=} \min_m E[U(Y-m)] = \min_m \int U(y-m)\cdot\rho(y)\,dy.$$

- If $Y \approx c_0 + c_1 \cdot X$, then using this estimate instead of m decreases disutility to:

$$d_U(Y\,|\,X) = \min_{c_0,c_1} \int U(y-(c_0+c_1\cdot x))\cdot\rho(x,y)\,dx\,dy.$$

- The corresponding decrease $D_U(Y\,|\,X)$ in disutility can be thus estimated as

$$D_U(Y\,|\,X) \stackrel{\text{def}}{=} \frac{d_U(Y)-d_U(Y\,|\,X)}{d(Y)}.$$

- When $U(d) = d^2$, we get mean and Pearson's correlation: $m = E[Y]$ and $D_U(Y\,|\,X) = R^2 = (r(F))^2$.

11. Approximate Linear Dependence (cont-d)

- In the above definitions, we only got $(r(F))^2$, i.e., only the *absolute value* $|r(F)|$ of the correlation $r(F)$.
- To get the *sign* of the correlation, we must separately consider $\uparrow$ and $\downarrow$ linear dependencies $c_0 + c_1 \cdot X$:

$$d^+_U(Y\,|\,X) = \min_{c_0;c_1\ge 0} \int U(y-(c_0+c_1\cdot x))\cdot \rho(x,y)\,dx\,dy;$$

$$d^-_U(Y\,|\,X) = \min_{c_0;c_1\le 0} \int U(y-(c_0+c_1\cdot x))\cdot \rho(x,y)\,dx\,dy;$$

$$\text{and } D^\pm_U(Y\,|\,X) \stackrel{\text{def}}{=} \frac{d_U(Y) - d^\pm_U(Y\,|\,X)}{d(Y)}.$$

- When $U(d) = d^2$, then:
  - $(r(F))^2 = D^+_U(Y\,|\,X)$ when $r(F) \ge 0$, and
  - $(r(F))^2 = D^-_U(Y\,|\,X)$ when $r(F) \le 0$.

12. Acknowledgments

This work was supported in part:

- by the National Science Foundation grants:
  - HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and
  - DUE-0926721,
- by Grants 1 T36 GM078000-01 and 1R43TR000173-01 from the National Institutes of Health, and
- by a grant N62909-12-1-7039 from the Office of Naval Research.